

# Technology Mapping for Low Leakage Power and High Speed with Hot-Carrier Effect Consideration<sup>1</sup>

Chang-woo Kang and Massoud Pedram

Dept. of Electrical Engineering-Systems

University of Southern California, Los Angeles CA 90089

{ckang,pedram}@usc.edu

**ABSTRACT** - Leakage power and hot-carrier effects are emerging as key concerns in deep sub-micron CMOS technologies with respect to their effects on the total power dissipation and reliability of VLSI circuits. Leakage power dissipation is rapidly becoming a substantial contributor to the total power dissipation as threshold voltage becomes small. Similarly, the hot-carrier effect is one of the most significant failure mechanisms in high-density VLSI circuits. In this paper, a technology mapping technique is presented for use in reducing the leakage power dissipation of the circuit by utilizing a dual-threshold voltage cell library and for minimizing the aged delay of the circuit by considering the effect of hot carriers on the cell speeds as the circuit ages. In addition, this paper presents two methods to reduce delay during technology mapping: primary output ordering and pin permutation. Experimental results show that the total power dissipation and leakage power dissipation can be reduced by up to 27% and 52% as a result of the leakage-aware technology mapping and that the circuit aging phenomenon can be reduced by up to 10.6% as a result of hot-carrier-aware technology mapping. Delay was also reduced by up to 13% by using primary output ordering and pin permutation.

## 1. INTRODUCTION

As device dimensions shrink to the deep sub-micron ranges, leakage power dissipation and the hot-carrier effect will become major issues and challenges for low-power and reliable systems in the near future.

Leakage power dissipation will become a substantial component as threshold voltages and channel lengths are reduced. Lowering supply voltage leads to reducing threshold voltage so as to regain performance at the cost of exponentially increasing leakage power [1]. There have been several approaches to reduce leakage power dissipation in standby mode. By turning transistors off in a stack, leakage current can be reduced considerably [1, 2]. This is called the stack effect. In order to utilize this effect, transistors were inserted to force transistors to be connected in series and to be turned off in standby mode [3]. In [1], a heuristic algorithm was proposed to find the best input that turns off the maximum number of devices in the transistor stacks in a circuit in order to minimize the leakage power dissipation during standby mode. As an alternative approach, dual-threshold voltage assignment schemes have been proposed [4-6]. In a dual-threshold process, an additional mask layer is used to assign either a high or low threshold voltage to each transistor. In [6], authors optimized the circuit speed and leakage power by using low-threshold and high-threshold voltage transistors in gates that lie on the timing critical and non-critical paths, respectively.

The hot-carrier-induced damage in MOS transistors is caused by the injection of high-energy electrons into the gate oxide near the drain region. Those injected carriers may be trapped in the oxide, which results in the degradation of the MOS transistor characteristics and can cause the degradation of the circuit [7]. Without considering

this fact, timing paths will change their delay characteristic and cause problems of system reliability. Techniques for estimating and reducing hot-carrier effects have been studied extensively [8, 9]. For example, Chang et al. [9] proposed rewiring, gate resizing, and pin reordering to reduce hot-carrier effects.

This paper presents a technology mapping technique that reduces total power dissipation, including leakage power dissipation, by utilizing both stack effect and a dual-threshold voltage library. Previous researchers have primarily focused on leakage power dissipation in standby mode. However, by considering leakage power dissipation in technology mapping, the leakage power dissipation can be reduced in the active mode as well as the standby mode. This proposed technique reduces the total power dissipation in the circuit while maintaining its logic speed. Furthermore, a simple aging model of a gate for hot-carrier effect is proposed and used for technology mapping. This model considers the delay degradation caused by hot-carrier effects so that it can optimize circuits for long-term reliability. On top of these schemes, two simple heuristic approaches are presented in which primary outputs of the circuit are ordered based on logic depth of their respective logic cones, and where gate input pins are permuted during the technology mapping algorithm. The objective of both heuristics is to reduce the circuit delay without impacting the total power dissipation.

The background work is discussed in Section 2. Two heuristic techniques to improve the circuit speed of mapped netlists are presented in Section 3. Models for leakage power and hot-carrier effect are proposed in Section 4. A technology mapping algorithm that captures the leakage power cost and the aging effect in transistors is described in Section 5. In Section 6, the power-speed trade-offs of technology mapping with a dual-threshold voltage library are discussed and results from heuristic speedup techniques are reported. Conclusions are given in Section 7.

## 2. BACKGROUND

This section briefly reviews power-delay technology mapping, threshold voltage scaling effects, and the hot-carrier effect.

### 2.1 Power-Delay Technology Mapping

Technology mapping problem can be described as follows: binding nodes in a Boolean network representing a combinational logic circuit optimized by technology-independent synthesis procedures to gates in a target library such that cost of the final implementation is minimized and timing constraints are satisfied.

In [10, 12], authors used a cost-delay curve to store all of the cost-delay trade-offs at a node to reduce the cost during the mapping phase. The cost is power dissipation in this paper. During the post-order traversal starting from primary inputs, a set of possible arrival times and power dissipations at each node of a network are produced. Starting at primary outputs, pre-order traversal determines

<sup>1</sup> This research was supported in part by DARPA PAC/C program under contract DAAB07-02-C-P302 and by NSF under grant no. 9988441.

the best match for each node, so that it can minimize power dissipation of the network while it satisfies the required time.

The post-order traversal can have two steps for each node: *adding* power-delay curves at the children of the node and then *merging* the curves. After adding and merging, a new curve must have only non-inferior points. A non-inferior point  $(t', p')$  is a point if and only if there does not exist a point  $(t, p)$  such that either  $t \leq t', p \leq p'$  or  $t < t', p \leq p'$ , where  $t$  is an arrival time and  $p$  represents power dissipation. For the pre-order traversal, the user is allowed to select the arrival time-power trade-off, which is the most suitable for his application. Given the required time at the root of the tree, a suitable point on the power-delay curve for the root node is chosen. The gate matching for the point at this root is identified and then required times for its inputs are computed. The preorder traversal resumes at its children nodes to satisfy the new required time and the minimum power dissipation.

On top of the mapping technique presented in [10], authors in [11] introduced an  $\alpha$ -approximate algorithm to compress a delay-cost function exponentially while keeping solutions within a constant bound of the optimal solution (cost may be area, power dissipation, etc.). The delay-cost function may use a large memory size as the size of library set increases. A simple solution limits the maximum number of points,  $K$ , which can be kept in any delay-cost curve [10]. Notice that if  $K$  is chosen to be too small, some optimal points may be dropped. On the other hand, if  $K$  is chosen to be too large, it will slow the mapping procedure. The authors of [11] introduced the notion of the  $i^{\text{th}}$   $\alpha$ -break point as defined by

$$\lambda_{\alpha}^i = c_i(1 + \alpha)^i \quad (1)$$

where  $(c_i, t_i)$  denotes the cheapest design point of a delay-cost curve and the cost might be area or power. The  $i^{\text{th}}$   $\alpha$ -interval is defined as the semi-open interval  $[\lambda_{\alpha}^i, \lambda_{\alpha}^{i+1})$ . The  $\alpha$ -delay-cost curve contains at most two points in every  $\alpha$ -interval; if a point in the curve belongs to some  $\alpha$ -interval, then it is either the cheapest or the costliest design solution in that interval. The maximum number of points in the  $\alpha$ -delay-cost curve is logarithmically proportional to the cost range  $[c_1, c_n]$ . The authors proved that the solution obtained by the exponential compression of the  $\alpha$ -delay-cost curve is within  $\alpha\%$  of the optimal solution for a tree-structured subject graph if all input pin capacitances for the library gates are the same. Thus, this technique – under equal pin capacitance assumption – provides an exponential compression compared to the original compression technique proposed in [10] and also used in [12].

## 2.2 Threshold Voltage Scaling

Demand for minimizing the dynamic power dissipation and push toward ever-shortening channel lengths in CMOS technology force the supply voltage to be scaled down. To maintain the circuit speed, the transistor threshold voltages must also be scaled down. This is easily seen from the first-order propagation delay equation of a transistor

$$\tau = \frac{CV_{dd}}{(V_{dd} - V_{TH})^{\alpha}} \quad (2)$$

where  $C$  is the load capacitance,  $V_{TH}$  is the threshold voltage, and  $\alpha$  (which is greater than 1 but less than 2) models the short channel effect [4]. The subthreshold leakage current can be modelled from the BSIM MOS transistor model as

$$I_{subth} = \mu_0 C_{ox} \frac{W_{eff}}{L_{eff}} \left( \frac{kT}{q} \right)^2 e^{1.8} e^{\frac{q}{kT}(V_{GS} - V_{TH0} - \gamma V_{SB} + \eta V_{DS})} \left( 1 - e^{-\frac{qV_{DS}}{kT}} \right) \quad (3)$$

where  $\mu_0$  is the zero bias mobility,  $C_{ox}$  is the gate oxide capacitance per unit area,  $kT/q$  is the thermal voltage,  $V_{TH0}$  is the zero bias

threshold voltage,  $\gamma$  is the linearized body effect coefficient, and  $\eta$  is the DIBL (Drain-induced barrier lowering) effect coefficient [1, 13]. Clearly, the leakage current increases exponentially as the threshold voltage is reduced.

## 2.3 Hot-Carrier Effect

As MOSFET devices are scaled down to small dimensions, hot-carriers injected into the MOS gate oxide layer cause a major reliability problem [14]. Trapping these carriers in the gate insulator layer causes degradation in the transistor transconductance and threshold voltage. Hot-carrier degradation takes place when a transistor is in the saturation region. CMOS circuits can be in this region during transitions. Therefore, a slow slew rate of input pins and a high load capacitance of an output pin of a gate will stress a transistor and a high switching activity in a gate will wear out the driving capability of a transistor quickly. In [15], authors proposed a ratio-based hot-carrier degradation model for the aging-aware timing simulation of large-scale circuits. The model utilized compact gate-level representation with timing information only, rather than a conventional complex transistor-level approach.

## 3. SPEEDUP HEURISTICS

In this section, two heuristic techniques to improve the circuit speed of mapped netlists are presented.

### 3.1 Primary Output Ordering

The technology mapping algorithm processes one primary output logic cone at a time. The order of mapping is important and produces different mapping results in terms of area, power dissipation, and circuit delay. To achieve better mapping solutions, a simple heuristic approach is adopted whereby the primary outputs of the given circuit are sorted in descending order of their (maximum) logic depths. Next, the logic cones are processed so that the most timing-critical cone is mapped first. The intuition is that the delay of a mapped cone is roughly proportional to the logic depth of the corresponding primitive gate cone. Furthermore, by processing cones with larger logic depths first, maximum flexibility is provided for mapping these timing critical logic cones. This simple and intuitive heuristic reduces the circuit power dissipation and delays will be shown in Section 5.

### 3.2 Pin Permutation

To reduce gate delay, it is well known that for sufficiently fast input transition times, the slowest input signal must be assigned to the top transistor of a pull-down section in a complex CMOS gate [20]. Namely, careful pin assignment for signals can result in reduced propagation delay through a CMOS gate. Therefore, during the technology mapping, this observation can be exploited in order to significantly reduce the overall delay of the mapped circuit netlist.

#### 3.2.1 Equivalent Pins

Each cell in the library is annotated with its pin swapping information, that is, pins of the cell are assigned to a number of pin sets such that all pins belonging to the same pin set can be interchanged during the technology mapping procedure. For example, if a cell implements Boolean function  $AB + C$ , the pin sets are  $\{A, B\}$  and  $\{C\}$ .

#### 3.2.2 Pin Permutation and Delay Calculation

During the technology mapping procedure, when a cell match is found at the output of some node in the subject network, all valid (as per pin set classification) input signal-to-pin assignments are enumerated, and the pin assignment that results in the lowest delay is selected to be included in the power-delay curve of the node in the subject graph. Consider the example of a four input NOR gate:

$$Y = \overline{(A+B+C+D)}$$

which has  $4! = 24$  valid permutations of the input pins. During mapping, a four-input NOR match and its best pin assignment are found and stored. The power-delay curve data is modified accordingly. The pin assignment information is recovered from the solutions stored in these curves during the output-to-input traversal that generates the complete mapping solution.

## 4. GATE MODELING

In this section, models for leakage power dissipation and gate aging due to the hot-carrier effect are proposed.

### 4.1 Power Dissipation

Power dissipation in a CMOS circuit consists of dynamic, short-circuit, internal, and leakage components. Charging and discharging the load capacitance of a gate causes dynamic power dissipation. Internal power dissipation results from charging and discharging the internal nodes of a gate without changing output value because of spurious input signals. Short-power dissipation appears when PMOS and NMOS transistors are on during output transitions. Leakage power dissipation exists because of the subthreshold current, which is dominant in current technology.

The average dynamic power dissipation of a gate in a synchronous CMOS circuit is given by

$$P_{dyn} = \frac{1}{2} \times C_{load} \times \frac{V_{dd}^2}{T_{cycle}} \times sw \quad (4)$$

where  $C_{load}$  is the load capacitance of the gate,  $V_{dd}$  is the supply voltage,  $T_{cycle}$  is the clock cycle time, and  $sw$  is the switching probability per cycle at the output of the gate[12]. The amount of leakage current is strongly dependent on the binary pattern applied to the inputs of a CMOS gate. This phenomenon is shown in Figure 1. Therefore, signal probability, which is defined as probability for a signal to assume a value of one, must be considered when estimating leakage power dissipation of a gate. The leakage power dissipation in a CMOS gate can be calculated as

$$P_{leak} = V_{dd} \times \sum_U [I_{subth}(U) \times pr(U)] \quad (5)$$

where  $U$  is an input vector for a gate,  $pr(U)$  is the probability that the input vector  $U$  is applied as an input pattern for the gate. For example, the probability for input vector  $U(0,1,1)$  of a three input gate, which has  $A, B$  and  $C$  as its input signals is calculated as

$$pr(U) = (1 - pr(A)) \times pr(B) \times pr(C). \quad (6)$$

Note that the signal probabilities of intermediate variables in a Boolean network can be calculated from the signal probabilities of the primary input of the circuit by using well-known signal probability propagation techniques.

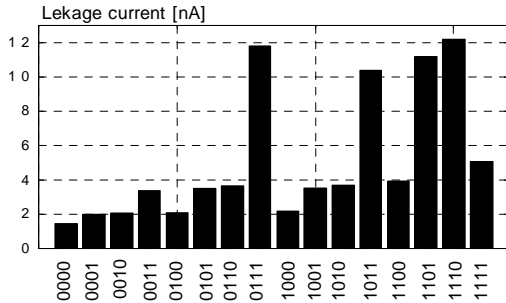


Figure 1: Leakage current of different input patterns for a four-input NAND gate. MSB refers to the input pin closest to the output port.

## 4.2 Gate Aging due to the Hot-Carrier Effect

The damage by the hot-carrier effect directly impacts the delay of a transistor because it shifts the threshold voltage and decreases the drain current driving capability [9]. A simple delay model is proposed for a gate to represent the aged delay due to the hot-carrier effect. The model can be described as follows:

$$T_{aged} = T_{fresh} \times \left( 1 + sw \times \delta C_{load} \frac{\zeta}{T_{avg.slew}} \right) \quad (7)$$

where  $T_{fresh}$  is a *fresh input-to-output delay*,  $sw$  is the switching probability of the gate output,  $\delta$  and  $\zeta$  are the degradation factors due to the output load and the slew rate of the input transition, respectively,  $C_{load}$  is the output load capacitance,  $T_{avg.slew}$  is an average input slew rate used for the purposes of normalization, and  $T_{aged}$  is the *aged input-to-output propagation delay*.  $\delta$  is inversely proportional to the driver strength of a gate, because the driver strength determines the duration of the transition time of a transistor.

In current semiconductor process technologies, hot-carrier induced degradation is much more severe in NMOS transistors than in PMOS [7]. Therefore, this paper only considers delay degradation for transistors in the pull-down network.

## 5. TECHNOLOGY MAPPING

Leakage power aware mapping with a dual-threshold voltage library and hot-carrier effect aware mapping for long-term performance optimization is presented here.

### 5.1 Leakage Power Aware Mapping (LPAM)

A low leakage power technology mapper has been implemented based on the algorithm presented in [12]. A leakage power model has been considered in the power dissipation, in order to choose the gate that dissipates the least amount of power at a node. The total power dissipation at a node  $n$  with a gate  $g$  is given by

$$P(n, g) = \frac{1}{2} C_{diff}(n, g) \frac{V_{dd}^2}{T_{cycle}} sw_n + V_{dd} \sum_U I_{subth}(U) pr(U) + \sum_{ni \in inputs(n, g)} \left( \frac{1}{2} C_{load}(ni) \frac{V_{dd}^2}{T_{cycle}} sw_{ni} + \frac{P(ni, gi)}{fanout(ni)} \right) \quad (8)$$

where  $C_{diff}(n, g)$  is the diffusion capacitance on an output port of a gate  $g$  matched on node  $n$ ,  $C_{load}(n)$  is the output load driven by node  $n$ ,  $sw_n$  is the switching probability of node  $n$ ,  $fanout(n)$  is the number of fanouts from node  $n$ , and  $P(ni, gi)$  is the average power dissipated at input  $i$ . Therefore, load capacitance is the gate input capacitance of a pin that a fanin node must drive.

### 5.2 Dual-Threshold Voltage Mapping

Several researchers [4-6, 16, 17] have proposed dual-threshold voltage assignment schemes for low leakage power during standby mode. They tried to assign low-threshold voltage on critical paths, while assigning high-threshold voltage on non-critical paths. These approaches have been performed with post-mapped circuits. However, this approach puts a serious constraint on the gates that can replace the current gate on a node, because only gates with the same Boolean function can be candidates for the replacement. In contrast, by having a library of cells with dual-threshold voltages, one can do a better job by choosing the best matching gate (with appropriate threshold voltage) during the technology-mapping phase. By performing threshold voltage assignment during the technology mapping phase, one will have much more flexibility in choosing the best set of logic gates and threshold voltages to meet the timing constraint with least power consumption.

### 5.3 Hot-Carrier Effect Aware Mapping

As in [12], the pin-dependent SIS library delay model is adopted here for calculation of the arrival time. However, the delay

calculation equation is modified to account for the aging effect due to the hot-carrier phenomenon as described below.

Suppose that gate  $g$  has matched at node  $n$ , and then the fresh pin-to-pin delay becomes:

$$T_{fresh}(n, g, C_{load}) = \tau_{i,g} + R_{i,g} C_{load} \quad (9)$$

where  $\tau_{i,g}$  is the intrinsic gate delay from input  $i$  to output of  $g$ ,  $R_{i,g}$  is drive resistance of  $g$  corresponding to a signal transition at input  $i$ , and  $C_n$  is the load capacitance seen at  $n$ . Arrival time is calculated as:

$$arrival(n, g, C_n) = \max_{ni \in inputs(n,g)} (T_{aged} + arrival(ni, gi, C_i)) \quad (10)$$

where  $T_{aged}$  is the aged time of node  $n$  given by equation (7),  $arrival(n, g, C_i)$  is the arrival time at input  $i$  corresponding to load  $C_i$  seen at that input,  $sw_n$  and  $T_{avg.slew}$  denote the switching probability of node  $n$  and average slew rate of input pins of gate  $g$ , and  $gi$  is the best match found at input  $i$ .

## 6. SIMULATION RESULTS

In this section, the simulation results for leakage power and for hot-carrier effect are described separately. SIS [18] was used for the synthesis environment and a 0.18  $\mu\text{m}$  CMOS technology cell library was employed, which is based on a 0.4V threshold voltage. Calculation of signal probabilities using ordered binary decision diagram (OBDDs) requires large amount of memory and long computation time. Therefore, we used the *local OBDD-based approach* presented in [23]. In that scheme, nodes in the network are first leveled. Then the OBDD variables for each local OBDD (associated with some node  $n$ ) are selected from the transitive fanins of  $n$  that are at least  $l$  levels away from  $n$ , where  $l$  is a user-supplied number. Please refer to [23] for additional details.

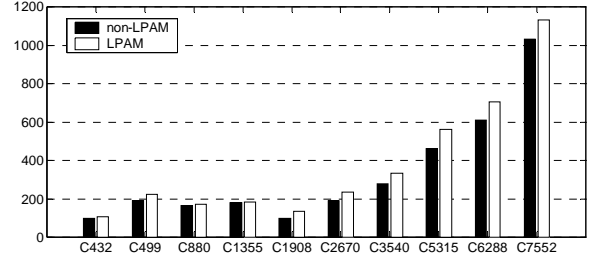
### 6.1 Leakage Power Reduction

To validate this scheme, a set of library cells with 0.2V threshold voltage was generated based on the original library of cells. For each gate in the two libraries, leakage current was obtained and recorded for every input pattern. Hspice simulations obtained the leakage currents. The circuits were simulated at 1.8-V supply voltage, and the primary input switching activity was set to 0.5. The circuit temperature was assumed to be 110  $^\circ\text{C}$  in the active mode.

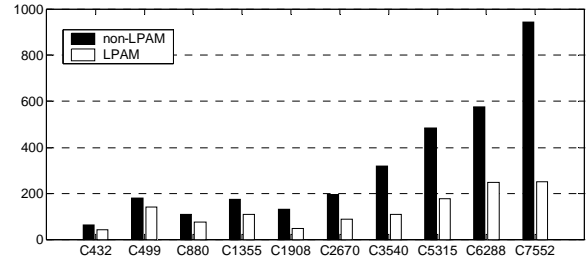
Figure 2 shows that the proposed technology mapping procedure reduces leakage power dissipation, resulting in a reduction of total power dissipation. Only low threshold voltage gates were used to measure the effectiveness of the leakage-aware mapping with the same delay constraints. As can be seen, leakage power is reduced by as much as 50%, while dynamic power increases by only 16%. As a result, 17% total power reduction was achieved. Therefore, although the dynamic power dissipation may increase after leakage optimization, because the technology mapper uses total power consumption (summation of the leakage and dynamic components) as its objective function, it will make the right decision. Furthermore, as the circuit size increases, the average activity per node tends to decrease whereas the total leakage increases. Therefore, the LPAM technique becomes even more effective as shown clearly in the experimental results of these tables.

There are several advantages to simultaneously mapping high-threshold voltage cells on non-critical paths and low-threshold voltage cells on critical paths. Leakage power is reduced substantially. In Figure 3, leakage power is reduced by 68% on average compared to that dissipated in circuits mapped with low-threshold voltage. Dynamic power dissipation has been decreased by 5%, because the gate capacitance of a transistor increases as its threshold voltage is lowered [21]. In fact, dynamic power can be reduced due to the reduction of internal-node voltage swing for high-

threshold cells [7]. In addition, as a part of dynamic power dissipation, short circuit power dissipation decreases as threshold voltage increases [22]. However, as the timing constraint becomes tight, dynamic power will increase to meet the constraint such as C7552 in Figure 7(a). These results led to 24% reduction in total power dissipation. When standby modes in portable devices are considered, this effect will be more even substantial.

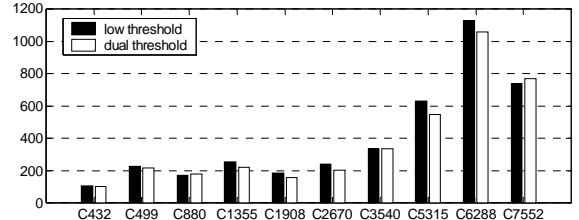


(a)

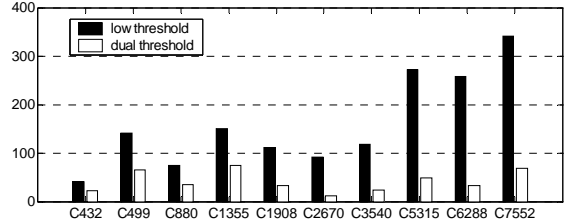


(b)

Figure 2: Power dissipation reduction by LPAM (leakage power aware mapping): (a) dynamic power dissipation in  $\mu\text{W}$ , (b) leakage power dissipation in  $\mu\text{W}$ .



(a)



(b)

Figure 3: Power dissipation reduction due to technology mapping with dual threshold cells: (a) dynamic power dissipation in  $\mu\text{W}$ , (b) leakage power dissipation in  $\mu\text{W}$ .

Based on our simulations, about 73 % of the gates can be mapped with high-threshold voltage cells on non-critical paths. There is also an advantage of maintaining performance. Note that technology

mapping was repeatedly performed to achieve the best circuit speed by starting with a loose timing constraint and gradually tightening it until the circuit could not meet the timing constraint. This tends to guarantee the high circuit speed while it reduces power dissipation. This is shown in Figure 4. By mapping low-threshold voltage cells on critical paths, circuits regained most of the circuit speed that was achieved when using low-threshold voltage cells everywhere. Finally, the area becomes larger than that of circuits mapped by low-threshold voltage only, because low threshold transistors have high driving strength with the same size of high threshold transistors. The result is shown in Figure 5. Those areas are to achieve approximately the same delay by using different library sets.

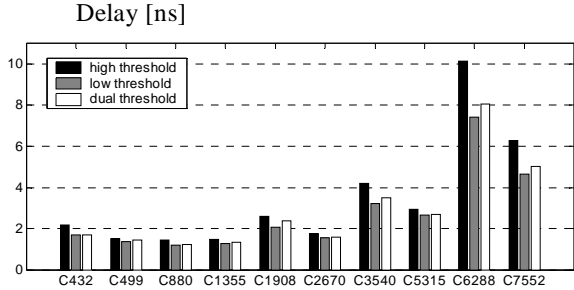


Figure 4: Speed-up by mapping low threshold voltage gates on critical paths.

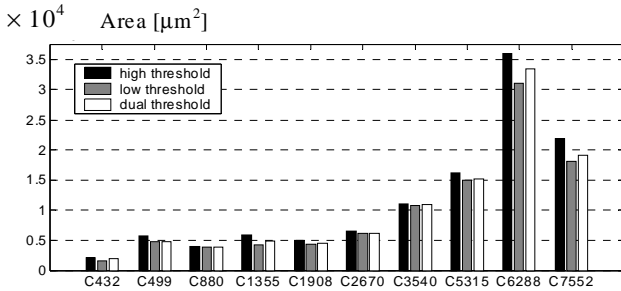


Figure 5: Area reduction by mapping a high-threshold voltage gate on non-critical paths.

## 6.2 Considering the Gate Aging Effect

Table 1 shows the experimental results for the MCNC benchmark circuits. As shown by these results, by considering the aging effect during the technology mapping phase, the aging-aware delay calculator can identify the logic paths that are likely to become critical after some aging and pass this information on to the mapper who will then generate the appropriate mapping solution that will ensure that the circuit will operate within the required timing constraints even after some degree of aging. In most cases, fresh delays from aging-aware mapping are also shorter than delays resulted from aging-unaware mapping, because the aging-aware mapper tries to meet the timing constraint with aged delays, which are longer than fresh delays.

## 6.3 Speed-up by Output Ordering and Pin Permutation

Table 2 shows the results of speed-up by primary output ordering based on the logic depth of circuits during technology mapping. The same timing constraint was given for two different sets of experiments. As indicated in the table, the delay improvement is up to 9.2%. Furthermore, one can see that power and area were also reduced except in lone case (i7). This general trend can be explained

as follows. Logic cones with larger depth tend to have a larger number of nodes in the NAND-decomposed network, and thus, tend to account for a larger portion of the total circuit area and power dissipation in the mapped netlist. By performing technology mapping on the larger cones first, we provide maximum flexibility to the technology mapper.

In some circuits, the delay can also be reduced by as much as 12.8% by performing pin permutation during the dynamic programming-based mapping as shown in Table 3. Notice that in Table 3 the delay improvement for some circuits comes at the expense of an area and power increase (see for example, C432 results). This may appear strange since, at least on the surface, pin permutation should not have increased the circuit area. The key to understanding these results is that because the pin permutation is integrated with the dynamic programming-based technology mapping algorithm, it potentially changes the area-delay curve at each node in the subject graph, and hence, at the end of the dynamic programming algorithm, the mapping solution will be quite different from the solution that is obtained without dynamic pin permutation. Indeed, if we performed the pin permutation step on a mapped circuit, then only its delay will be impacted while its area will remain exactly the same. Notice however that in such a case, the circuit delay reduction will be less than when pin permutation is intrinsically integrated within the dynamic programming algorithm as we have proposed and implemented here.

Table 1: Fresh and aged delays by aging-unaware and aging-aware mapper.

Circuit	Aging-unaware mapping		Aging-aware mapping		Speed-up	
	Fresh [ns]	Aged [ns]	Fresh [ns]	Aged [ns]	Fresh (%)	Aged (%)
C432	2.21	2.54	2.08	2.34	6.3	7.9
C499	1.52	1.74	1.5	1.66	1.3	4.6
C880	1.39	1.41	1.33	1.34	4.5	5.0
C1355	1.61	2.09	1.72	1.94	-6.4	7.2
C1908	2.65	2.82	2.45	2.52	8.2	10.6
C2670	1.89	1.93	1.81	1.86	4.4	3.6
sqrt8ml	1.93	2.06	1.97	1.98	-2.0	3.9
f51m	2.01	2.04	1.89	1.95	6.3	4.4
alu2	2.1	2.14	2.03	2.05	3.4	4.2
i7	1.13	1.14	1.1	1.11	2.7	2.6

## 7. CONCLUSION

In this paper, a technology mapping technique was presented to reduce leakage power dissipation as well as total power dissipation. By considering leakage power dissipation based on input signal probabilities, the reduction on total power dissipation became substantial as circuit size increased. We also presented the trade-offs of mapping dual-threshold voltage library with respect to leakage power dissipation, total power dissipation, delay, and area. The mapper mapped nodes on non-critical paths with high-threshold-voltage gates, resulting in an up to 52% reduction in leakage power dissipation and 27% in total power dissipation. In addition to the leakage power optimization, an aging model was proposed for the technology mapping to represent the transistor degradation due to the hot-carrier effect. The aging phenomenon was reduced by up to 10.6% in the test benchmark circuits. Two different methods for improving delay during technology mapping were presented: primary output ordering and pin permutation. They showed up to about 9% speed-up compared to results without those schemes.

Table 2: Technology mapping with primary output ordering

Circuit	W/o PO ordering			W/ PO ordering			% Improvement		
	Area ( $\mu\text{m}^2$ )	Power ( $\mu\text{W}$ )	Delay (ns)	Area ( $\mu\text{m}^2$ )	Power ( $\mu\text{W}$ )	Delay (ns)	Area	Power	Delay
C432	2219	82.0	2.48	2128	78.59	2.36	4.3	4.2	4.9
C499	3881	136.9	1.65	3857	136.3	1.62	0.6	0.4	1.8
C880	4232	170.4	1.84	3986	155.7	1.67	6.2	8.6	9.2
C1355	6390	348.5	2.04	5988	330.7	2	6.7	5.1	2
f51m	1269	56	2.09	1255	54	2.05	1.1	3.6	1.9
alu2	4283	131.1	2.93	4270	125.4	2.84	0.3	4.3	0.3
i7	5271	142.4	1.51	5278	141.1	1.5	-0.1	0.8	0.7

Table 3: Technology mapping with pin permutation

Circuit	W/o Pin Permutation			W/ Pin Permutation			% Improvement		
	Area ( $\mu\text{m}^2$ )	Power ( $\mu\text{W}$ )	Delay (ns)	Area ( $\mu\text{m}^2$ )	Power ( $\mu\text{W}$ )	Delay (ns)	Area	Power	Delay
C432	3207	101.5	2.27	3331	103.5	2.04	-3.9	-2	10
C499	3395	125.8	1.53	3380	126	1.49	0.4	0	2.6
C880	2637	155.1	1.90	2865	162	1.7	-8.6	-4.5	10.5
C1355	3428	370.8	1.61	3380	377.1	1.54	1.4	-1.7	4.3
f51m	1026	79.82	1.81	997	73.04	1.71	2.9	8.5	5.5
alu2	2950	148.3	2.58	2958	147.1	2.43	-0.3	0.8	5.8
i7	4128	109.1	1.64	4134	107.9	1.43	-0.2	1.1	12.8

## REFERENCES

- [1] Chen, Z. and I. Koren, "Technology mapping for hot-carrier reliability enhancement," in *Proc. of the SPIE - The International Society for Optical Engineering*, pp. 42-50, 1997.
- [2] Johnson, M.C., et al., "Models and algorithms for bounds on leakage in CMOS circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 18, pp. 714-725, 1999.
- [3] Gu, R.X. and M.I. Elmasry, "Power dissipation analysis and optimization of deep submicron CMOS digital circuits," *IEEE Journal of Solid-State Circuits*, vol. 31, pp. 707-713, 1996.
- [4] Johnson, M.C., et al., "Leakage control with efficient use of transistor stacks in single threshold CMOS," in *Proc. of the Design Automation Conference*, pp. 442-445, 1999.
- [5] Kao, J.T. and A.P. Chandrakasan, "Dual-threshold voltage techniques for low-power digital circuits," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 1009-1017, 2000.
- [6] Pant, P., et al., "Dual-threshold voltage assignment with transistor sizing for low power CMOS circuits," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 9, pp. 390-394, 2001.
- [7] Wei, L., et al., "Design and optimization of dual-threshold circuits for low-voltage low-power applications," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 7, pp. 16-24, 1999.
- [8] Leblebici, Y., "Design consideration for CMOS digital circuits with improved hot-carrier reliability," *IEEE Journal of Solid State Circuits*, vol. 31, pp. 1014-1024, 1996.
- [9] Chang, C-W., et al., "Layout-driven hot-carrier degradation minimization using logic restructuring techniques," in *Proc. of the Design Automation Conference*, pp. 97-101, 2001.
- [10] Chaudhary K. and M. Pedram, "Computing the area versus delay trade-off curves in technology mapping," *IEEE Trans. on Computer Aided Design*, vol. 14, no. 12, pp. 1480-1489, 1995.
- [11] Roy, S., et al., "An alpha-approximate algorithm for delay -constraint technology mapping," in *Proc. of the Design Automation Conference*, pp. 367-371, 1999.
- [12] Tsui, C.-Y., et al., "Technology decomposition and mapping targeting low power dissipation," in *Proc. of the Design Automation Conference*, pp. 68-73, 1993.
- [13] Sheu, B.J., et al., "BSIM: Berkeley short-channel IGFET model for MOS transistors," *IEEE Journal of Solid State Circuits*, vol. 22, pp. 558-566, 1987.
- [14] Roy, K. and S.C. Prasad, *Low-power CMOS VLSI Circuit Design*. Wiley-Interscience, 2000.
- [15] Yonezawa, H., et al., "Ratio based hot-carrier degradation for aged timing simulation of millions of transistors digital circuits," *IEEE Int. Electron Devices Meeting Technical Digest*, vol. pp. 93-96, 1998.
- [16] Sundararajan, V. and D.K. Parhi, "Low power synthesis of dual threshold voltage CMOS VLSI circuits," in *Proc. of the International Symposium on Low Power Electronics and Design*, pp. 139-144, 1999.
- [17] Tripathi, N., et al., "Optimal assignment of high threshold voltage for synthesizing dual threshold CMOS circuits," in *Proc. of the International Conference on VLSI Design*, pp. 227-232, 2000.
- [18] Sentovich, E.M., et al., *SIS: A system for sequential circuit synthesis*, 1992, ERL, University of California, Berkeley.
- [19] Pedram, M., "Power minimization in IC design: principles and applications," *ACM Transactions on Design Automation of Electronics Systems*, vol. 1, no. 1, pp. 3-56, 1996.
- [20] J. M. Rabaey, *Digital integrated circuits: a design perspective*, Upper Saddle River, NJ: Prentice Hall, pp. 198 - 199, 1996.
- [21] S. Sirichotiyakul, et al., "Stand-by power minimization through simultaneous threshold voltage selection and circuit sizing," in *Proc. the Design Automation Conference*, pp. 436-441, 1999.
- [22] Veendrick, H.J., "Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits," *IEEE Journal of Solid-State Circuits*, vol. sc-19, pp. 468-473, 1984.
- [23] Ding, C-S, et al., "Gate-level power estimation using tagged probabilistic simulation," *IEEE Trans. on Computer Aided Design*, vol. 17, no. 11, pp.1099-1107, 1984.